54

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/981,996	10/19/2001	Yasunaga Miyazawa	110824	7093

7590 08/03/2005

OLIFF & BERRIDGE, PLC
P.O. Box 19928
Alexandria, VA 22320

EXAMINER	
SKED, MATTHEW J	

ART UNIT	PAPER NUMBER
2655	

DATE MAILED: 08/03/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)
	09/981,996	MIYAZAWA, YASUNAGA
	Examiner	**Art Unit**
	Matthew J. Sked	2655

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on 19 May 2005.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) 1-42 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 1-42 is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application (PTO-152)

6) ☐ Other: _____.

DETAILED ACTION

Response to Amendment

1. Applicant's arguments with respect to claims 1-33 have been considered but are moot in view of the new ground(s) of rejection.

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

3. Claims 1, 2, 4-5, 9, 10, 12, 13, 15-16, 20, 21, 23, 24, 26, 27, 31, 32, 34, 36, 37, 39, 40 and 42 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wakisaka (U.S. Pat 6,148,105) in view of Wymore (U.S. Pat 6,631,348).

As per claims 1, 12 and 23, Wakisaka teaches a speech recognition system comprising:

creating speech data on which different types of noise have been superposed (voice contains noise, col. 13, lines 1-3);

creating and storing acoustic models according to each of the noise types (creates multiple acoustic models under different noise environments, col. 14, lines 35-39);

during speech recognition: determining the type of noise superposed on speech data to be recognized (recognizes the type of noise, col. 13, lines 41-45);

selecting the corresponding acoustic model corresponding to the determined noise type (determined noise environment is used to search for an acoustic mode for voice recognition hence selecting a model, col. 13, lines 41-48);

eliminating the noise using a predetermined noise elimination method in the training and speech recognition processes (noise deletion unit, col. 13, lines 29-34); and

perform speech recognition based on the selected model (acoustic collating unit collates the input voiced with an acoustic model and produces a recognition result, col. 13, line 63 to col. 14, line 3).

Wakisaka does not teach the acoustic models corresponding to each of the noise types also contain a plurality of S/N ratios for each noise type.

Wymore teaches acoustic models corresponding to a plurality of S/N ratios (col. 4, lines 18-27). Specifically Wymore teaches different noise levels which implies the noise levels having different SNRs.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the acoustic models corresponding to noise types of Wakisaka to include the noise levels as taught by Wymore because, as taught by Wymore, moving from a serene environment to an environment with a high level of noise without changing S/N models would decrease accuracy (col. 2, lines 4-20).

4. As per claims 2, 13 and 24, Wakisaka teaches the speech recognition method according to claim 1, wherein the noise elimination method is at least one of a spectral subtraction method and a continuous spectral subtraction method, and the acoustic models are created by eliminating the noise by the at least one of the spectral

subtraction method and the continuous spectral subtraction method from each of the speech data on which the different types of noise have been superposed, obtaining the feature vectors of each of the speech data which have undergone the noise elimination, and using the feature vectors;

when speech recognition is performed, a first speech feature analysis is performed to obtain frequency-domain feature data of the speech data on which the noise has been superposed (obtains the spectrum for spectral subtraction from the inputted announcement voice, col. 15, lines 43-46);

a determination is made whether the speech data is a noise segment or a speech segment based on the result of the feature analysis, and when a noise segment is detected, the feature data thereof is stored (stores in the database data of the overall sound when there is no voice hence a determination is made if a speech segment is present, col. 14, lines 7-12), whereas when a speech segment is detected, the type of the noise superposed is determined based on the feature data having been stored and a corresponding acoustic model is selected from the acoustic models corresponding to each of the noise types based on the result of the determination (determined noise environment is used to search for an acoustic mode for voice recognition hence selecting a model, col. 13, lines 41-48);

the noise is eliminated by the at least one of the spectral subtraction method and the continuous spectral subtraction method from the speech data to be recognized on which the noise has been superposed (performs spectral subtraction, col. 15, lines 43-46); and

a second feature analysis is performed on the speech data which has undergone the noise elimination to obtain feature data required in the speech recognition (sound analysis unit performs feature extraction processing on the noiseless voice, col. 13, lines 57-61) and a speech recognition is performed on the result of the feature analysis based on the selected acoustic model (acoustic collating unit collates the input voiced with an acoustic model and produces a recognition result, col. 13, line 63 to col. 14, line 3).

5. As per claims 4, 15, and 26, Wakisaka teaches eliminating the noises from each of the speech data by a predetermined noise elimination method, and using the feature vectors of each of the speech data which have undergone the noise elimination (stores data of voice that was obtained by removing noise, col. 14, lines 7-12).

Wakisaka does not teach the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created by generating speech data on which noises with the plurality of S/N ratios for each of the noise types have been respectively superposed.

Wymore teaches the acoustic models corresponding to the plurality of S/N ratios are created by generating speech data on which noises with the plurality of S/N ratios for each of the noise types have been respectively superposed and generates reference patterns for the multiple noise levels based on the training information (col. 4, lines 1-36).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the system of Wakisaka to create acoustic models for a plurality of

S/N ratios from training data as taught by Wymore because using a semi-clean speech model to recognize speech would give better recognition results than using only a noisy speech model.

6. As per claims 5, 16, and 27, Wakisaka does not teach when the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created, in addition to determining the type of the noise superposed on the speech data to be recognized, the S/N ratio is obtained from a magnitude of the noise in a noise segment and a magnitude of the speech in a speech segment, and an acoustic model is selected based on the S/N ratio obtained.

Wymore does not explicitly teach estimating the S/N ratio from the magnitude of the noise in the noise and the magnitude of speech in the speech segment, but he teaches choosing the acoustic model based on S/N ratios (noise levels, col. 4, lines 48-52).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the system of Wakisaka to obtain the S/N ratio and select an acoustic model based on both the noise type and S/N ratio as taught by Wymore because it would give a more robust estimate of the S/N ratio and hence giving a better estimate to choose a most appropriate acoustic model.

7. As per claims 9, 20 and 31, Wakisaka teaches a speech recognition system comprising:

creating speech data on which different types of noise have been superposed (voice contains noise, col. 13, lines 1-3);

eliminating the noise using a predetermined noise elimination method in the training and speech recognition processes (noise deletion unit, col. 13, lines 29-34);

creating and storing acoustic models according to each of the noise types (creates multiple acoustic models under different noise environments, col. 14, lines 35-39);

during speech recognition: determining the type of noise superposed on speech data to be recognized (recognizes the type of noise, col. 13, lines 41-45);

selecting the corresponding acoustic model corresponding to the determined noise type (determined noise environment is used to search for an acoustic mode for voice recognition hence selecting a model, col. 13, lines 41-48);

perform speech recognition based on the selected model (acoustic collating unit collates the input voiced with an acoustic model and produces a recognition result, col. 13, line 63 to col. 14, line 3),

wherein, when speech data on which another type of noise has been superposed is created, other acoustic models are created corresponding to the other noise type (creates acoustic models from noise and noiseless voice data, col. 14, lines 7-23).

Wakisaka does not teach the acoustic models corresponding to each of the noise types also contain a plurality of S/N ratios for each noise type.

Wymore teaches acoustic models corresponding to a plurality of S/N ratios (col. 4, lines 18-27). Specifically Wymore teaches different noise levels which implies the noise levels having different SNRs.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the acoustic models corresponding to noise types of Wakisaka to include the noise levels as taught by Wymore because, as taught by Wymore, moving from a serene environment to an environment with a high level of noise without changing S/N models would decrease accuracy (col. 2, lines 4-20).

8. Regarding claims 10, 21, and 32, Wakisaka teaches a system for reducing noise in speech signal that uses spectral subtraction (col. 15, lines 43-50).

9. As per claim 34, 36, 37, 39, 40 and 42, neither Wakisaka nor Wymore specifically teach the total number of acoustic models equals N x L, where N is a number of different noise types, and L is a number of S/N ratios for each of the noise types.

However, the obvious combination of using N multiple noise type models taught by Wakisaka and the L multiple noise level models taught by Wymore would necessarily produce the total number of acoustic models being the product of noise type models and noise level models.

10. Claims 3, 6-8, 11, 14, 17-19, 22, 25, 28-30, 33, 35, 38 and 41 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wakisaka in view of Wymore as applied to claims 1, 12, and 23 above, and taken in further view of Takagi (U.S. Pat 5,890,113).

As per claims 3, 14, and 25, Wakisaka teaches storing feature data when noise is detected (no announcement voice, col. 14, lines 7-12) and storing feature data when speech is detected (stores data from voice obtained by removing noise, col. 14, lines 7-12).

Wakisaka and Wymore do not teach a feature analysis to obtain a vector of cepstrum coefficients for use in detecting noise.

Takagi teaches extracting cepstrum coefficients from the sequence for speech recognition (analyzing unit, col. 7, lines 15-20) and uses it to detect noise (speech parts, col. 7, lines 44-46).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the system of Wakisaka and Wymore to extract cepstral features from the signal to be recognized as taught by Takagi because cepstral coefficients have highly desirable properties for speech recognition and classification.

Wakisaka and Wymore do not teach using cepstrum mean normalization method in noise elimination.

Takagi teaches using the cepstrum mean normalization to extract noise from an inputted speech signal in a speech recognition system (environmental adapting unit, col. 7, lines 37-40).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the system of Wymore and Wakisaka to eliminate the noise through cepstrum mean normalization because it is known to be a useful approach used for compensating for multiple distortions in a speech signal.

11. As per claims, 6, 17, and 28, Wakisaka teaches a speech recognition system comprising:

creating speech data on which different types of noise have been superposed (voice contains noise, col. 13, lines 1-3);

the noise is eliminated by the at least one of the spectral subtraction method and the continuous spectral subtraction method from the speech data to be recognized on which the noise has been superposed (performs spectral subtraction, col. 15, lines 43-46);

creating and storing acoustic models according to each of the noise types (creates multiple acoustic models under different noise environments, col. 14, lines 35-39);

when speech recognition is performed, a first speech feature analysis is performed to obtain frequency-domain feature data of the speech data on which the noise has been superposed (obtains the spectrum for spectral subtraction from the inputted announcement voice, col. 15, lines 43-46);

a determination is made whether the speech data is a noise segment or a speech segment based on the result of the feature analysis, and when a noise segment is detected, the feature data thereof is stored (stores in the data base data of the overall sound when there is no voice hence a determination is made if a speech segment is present, col. 14, lines 7-12), and when a speech segment is detected, the noise is eliminated from the speech segment by the spectral subtraction method or the continuous spectral subtraction method (performs spectral subtraction, col. 15, lines 43-46);

a second feature analysis is performed on the speech data which has undergone the noise elimination to obtain feature data required in the speech recognition (sound

analysis unit performs feature extraction processing on the noiseless voice, col. 13, lines 57-61);

when the speech segment has terminated, the type of the noise superposed is determined based on the feature data of the noise segment having been stored, and an acoustic model is selected from the acoustic models corresponding to each of the noise types (determined noise environment is used to search for an acoustic mode for voice recognition hence selecting a model, col. 13, lines 41-48); and

a speech recognition is performed on the result of the feature analysis based on the selected acoustic model (acoustic collating unit collates the input voiced with an acoustic model and produces a recognition result, col. 13, line 63 to col. 14, line 3).

Wakisaka does not teach the acoustic models corresponding to each of the noise types also contain a plurality of S/N ratios for each noise type.

Wymore teaches acoustic models corresponding to a plurality of S/N ratios (col. 4, lines 18-27). Specifically Wymore teaches different noise levels which implies the noise levels having different SNRs.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the acoustic models corresponding to noise types of Wakisaka to include the noise levels as taught by Wymore because, as taught by Wymore, moving from a serene environment to an environment with a high level of noise without changing S/N models would decrease accuracy (col. 2, lines 4-20).

Wakisaka and Wymore do not teach using a cepstrum mean normalization method to obtain feature vectors.

Takagi teaches using the cepstrum mean normalization to extract noise from an inputted speech signal in a speech recognition system (environmental adapting unit, col. 7, lines 37-40).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the system of Wakisaka and Wymore to eliminate the noise through cepstrum mean normalization as taught by Takagi because it is a common approach used for compensating for multiple distortions in a speech signal.

Wakisaka and Wymore do not explicitly teach using cepstrum coefficients in detecting noise.

Takagi teaches extracting cepstrum coefficients from the sequence for speech recognition (analyzing unit, col. 7, lines 15-20) and uses it to detect noise (speech parts, col. 7, lines 44-46).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the system of Wakisaka and Wymore to extract cepstral features from the signal to be recognized as taught by Takagi because cepstral coefficients have highly desirable properties for speech recognition and classification.

12. As per claims 7, 18, and 29, Wakisaka teaches eliminating the noises from each of the speech data by a predetermined noise elimination method, and using the feature vectors of each of the speech data which have undergone the noise elimination (stores data of voice that was obtained by removing noise, col. 14, lines 7-12).

Wakisaka does not teach the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created by generating speech data on which

noises with the plurality of S/N ratios for each of the noise types have been respectively superposed.

Wymore teaches the acoustic models corresponding to the plurality of S/N ratios are created by generating speech data on which noises with the plurality of S/N ratios for each of the noise types have been respectively superposed and generates reference patterns for the multiple noise levels based on the training information (col. 4, lines 1-36).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the system of Wakisaka to create acoustic models for a plurality of S/N ratios from training data as taught by Wymore because this would allow the models to be trained for different types and different levels of speech hence giving better speech recognition.

13. As per claims 8, 19, and 30, Wakisaka does not teach when the acoustic models corresponding to the plurality of S/N ratios for each of the noise types are created, in addition to determining the type of the noise superposed on the speech data to be recognized, the S/N ratio is obtained from a magnitude of the noise in a noise segment and a magnitude of the speech in a speech segment, and an acoustic model is selected based on the S/N ratio obtained.

Wymore does not explicitly teach estimating the S/N ratio from the magnitude of the noise in the noise and the magnitude of speech in the speech segment, but he teaches choosing the acoustic model on S/N ratios (noise levels, col. 4, lines 48-52).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the system of Wakisaka to obtain the S/N ratio and select an acoustic model based on both the noise type and S/N ratio as taught by Wymore because it would give a more robust estimate of the S/N ratio and hence giving a better estimate to choose a most appropriate acoustic model.

14. As per claims 11, 22 and 33, Wakisaka and Wymore do not teach using cepstrum mean normalization method in noise elimination.

Takagi teaches using the cepstrum mean normalization to extract noise from an inputted speech signal in a speech recognition system (environmental adapting unit, col. 7, lines 37-40).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify the system of Wakisaka and Wymore to eliminate the noise through cepstrum mean normalization because it is a well-known and convenient approach for compensating for multiple distortions in a speech signal.

15. As per claim 35, 38, and 41, Wakisaka, Wymore and Takagi do not specifically teach the total number of acoustic models equals N x L, where N is a number of different noise types, and L is a number of S/N ratios for each of the noise types. However, the obvious combination of using N multiple noise type models taught by Wakisaka and the L multiple noise level models taught by Wymore would necessarily produce the total number of acoustic models being the product of noise type models and noise level models.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Matthew J. Sked whose telephone number is (571) 272-7627. The examiner can normally be reached on Mon-Fri (8:00 am - 4:30 pm).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Wayne Young can be reached on 571-272-7582. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

MS
08/01/05

Susan McFadden
SUSAN MCFADDEN
PRIMARY EXAMINER